Ref. Ares(2021)4901578 - 01/08/2021

PROJECT DELIVERABLE REPORT

NAIADES

Greening the economy in line with the sustainable development goals

D4.9 AI-Water Quality Monitoring & Dynamical Water Treatment & DSS

A holistic water ecosystem for digitisation of urban water sector

SC5-11-2018

Digital solutions for water: linking the physical and digital world for water solutions

"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820985"

Document Information

Grant Agreement Number	820985	Acronym	NAIADES
Full Title	A holistic water ecosystem for digitization of urban water sector		
Topic	SC5-11-2018: Digital solutions for water: linking the physical and digital world for water solutions		
Funding scheme	IA - Innovation action		
Start Date	1st JUNE 2019	**Duration**	36 months
Project URL	www.NAIADES-project.eu		
EU Project Officer	Alexandre VACHER		
Project Coordinator	CENTER FOR RESEARCH AND TECHNOLOGY HELLAS - CERTH		
Deliverable	D4.9 AI-Water Quality Monitoring & Dynamical Water Treatment & DSS		
Work Package	WP4 – Smart Framework: AI monitoring, optimization and treatment modules		
Date of Delivery	**Contractual** M18	**Actual**	M17
Nature	R - Report	**Dissemination Level**	PU-PUBLIC
Lead Beneficiary	AIMEN		
Responsible Author	Dr. Juan Manuel Fernández Montenegro	Email	Juan.fernandez@aimen.es
		Phone	+34697 99 14 97
Reviewer(s):	UDGA, KT		
Keywords	Artificial Intelligence, machine learning, drinking Water Treatments Plan, DSS, treatment suggestions		

Revision History

Version	Date	Responsible	Description/Remarks/Reason for changes
0.1	09/10/2020	AIMEN	Table of Contents
0.2	09/11/2020	AIMEN	Report write-up
0.3	24/11/2020	UDGA, KT, AIMEN	Internal Review
0.4	26/11/2020	AIMEN	New suggestions by UDGA included
1.0	30/11/2020		Review and Release
1.1	11/05/2021	AIMEN	PO review applied
	19/05/2021	UDGA	Internal Review
1.2	19/05/2021	AIMEN	Review summarized in Section 7
2.0	27/05/2021	AIMEN	Review and Release

Contents

Abbreviations

ANN	Artificial Neural Networks
DCA	Data Collector and Aggregator
DNN	Deep Neural Network
DM	Data Manager
DSS	Decision Support System
DWTS	Dynamical Water Treatments Suggestions
ls-WTP	Laboratory scaled Water Treatment Plant
WTP	Water Treatment Plant

1 Summary

This report focuses on the description of the procedures and first prototype developed on Task 4.5 – AI-Water Quality Monitoring & Dynamical Water Treatment & DSS. This mid-term deliverable contains information that describes the purpose of the tool, the design and functioning of the tool and all the extra features that are required to be part of NAIADES platform. It also describes all the requirements necessary for the tool to perform positively, being the availability of quality real data and the related treatments the principal one. Additionally, it includes a description of the technical modules required to collect, aggregate and send data to the NAIADES cloud platform, that are being installed in the laboratory water treatment plant created in Task 4.2.

2 Introduction

Water from natural courses (rivers, lakes) is fully treated before being supplied to a distribution system from where it will go to consumers in drinking water supplies systems. Water treatments consists in sequential units to eliminate pollutants and pathogens. It usually includes pre-treatment, coagulation, flocculation and sedimentation, filtration and disinfection. Performance models can help in understanding and predicting treatments effectiveness, especially in stream events when abstraction water changes, like during storms or droughts. If these models have good accuracy, they could be used for treatment control in order to ensure that quality of the treated water mitigates the risks of potable water supply.

In NAIADES project, water treatment models will be used to help operators in water plants to monitor and control the process. This will be held in a decision support system that will make recommendations for better performance of the plant taking into consideration of water quality in real time at abstraction point.

The goal of this component is to be able to detect and react to the events related to the quality of water in a dynamical and reactive data driven approach, and to increase the performance and safety of the current systems. It will focus mainly on water quality changes occurred from extreme events such as heavy rains and droughts, thus being able to suggest treatments for five different events as described in NAIADES deliverable D4.3: normal operation, heavy rain, droughts, biological contamination and saline intrusion.

While mathematical models are usually used to describe and simulate different processes such as sedimentation, coagulation-flocculation, filtration, aeration, chemical oxidation or granular activated carbon adsorption that are used for water treatment, Artificial Intelligence, and in particular machine learning tools such as artificial neural networks (ANN) are robust technologies that can handle the complex and dynamic nature of water treatment processes.

This module implements an artificial intelligent support system, using state of the art machine learning techniques, to provide useful information to the experts on the selection and application of appropriated water treatments for the current scenario.

This document is structured as follows. Section 3 describes the proposed solution for NAIADES platform and for the specific end user problem. Section 4 describes the infrastructure required to collect data necessary for this solution. Section 5 focuses on the treatments suggestion tool to be deployed within NAIADES platform. Section 6 summarises the planned steps from M18 and Section 8 shows the conclusions.

3 NAIADES solutions

The proposed solution aims to provide suggestions about the optimal treatments (chemical dosages and processes times) adapted for any incoming water quality in order to ensure the best drinking water quality, especially during extreme events such as heavy rains or droughts. In order to achieve the following tasks are necessary to be tackled:

- Real-time water quality monitoring at the inlet of a drinking Water Treatment Plant (WTP), together with the treatments (dosages and times) applied to obtain optimal water quality at the outlet.
- Treatments simulation for data generation.
- Machine learning modelling.
- Decision Support System (DSS) for treatments recommendation.

The first two points are necessary to generate a model able to link the water quality parameters to the best treatments. This is done by NAIADES T4.2. The last two points will be handled by NAIADES Dynamical Water Treatments Suggestions (DWTS) Tool. This tool receives the quality parameters from the inlet of a known drinking WTP and returns suggestions about the best dosages and process times that will guarantee optimal drinking water. Any new WTP that want to use the suggestions tool should have the same treatment processes as the ones modelled by the tool, otherwise, it will require the tool to train a new model.

The tool will be providing real time suggestions and it could be easily combined with other NAIADES services to also provide future suggestions (advanced solution). An advanced solution could make use of NAIADES Water Quality Prediction Tool (see NAIADES Deliverable 5.9). This solution would make use of the historical inlet water quality values and weather forecasts to predict future inlet water quality (since the water quality is affected by weather events such as heavy rain increasing water turbidity), thus, the dynamical treatments suggestion tool can use those water quality inlet predictions to generate future suggestions. As a result, the DWTS Tool could be able to suggest real-time and also so many future suggestions as the predictions provided by the Water Quality Prediction Tool. The demonstration of this advanced solution is under constraints such as the necessity of historical data generation (water quality parameters with the correspondent weather ones); and difficulties on its application on the validation laboratory, so the methodology to apply it is being evaluated. Currently, all the requirements from NAIADES framework are being implemented (entities creation, data collection) but its demonstration will depend on the remaining time after completing the main functionality of this module (dynamic real time treatments suggestions).

3.1.1 Braila WTP based solution.

Braila was the end user interested in this solution but due to confidentiality issues, their drinking WTP could not provide all the required data for simulation, modelling and validation. The countermeasure plan was activated due to these circumstances, so it was planned the creation of a laboratory drinking WTP based on Braila's WTP infrastructure. This laboratory scaled WTP (ls-WTP) was initially conceived for model validation in extreme events during the last 6 months of the project.

The laboratory, currently under construction, will contain a scaled version of Braila's WTP treatments processes (coagulation, filtration and chlorination). This configuration is also very common in others WTP. Its inlet water quality parameters and treatments are going to be monitored real-time and different simulation will be undertaken as described in D4.3. The data collected during those simulations will be used to generate a model able to relate any water quality to the best treatments. The monitored inlet water quality parameters and the resultant suggestions will be then shared with NAIADES platform as a showcase, so Braila and any other WTP can validate the usability of this tool.

4 IoT system

This Section describes the IoT system required to collect data from the laboratory WTP all the way to NAIADES platform. The laboratory infrastructure as described in D4.3 will have tanks, vessels and pipes where the coagulation, filtration and chlorination will take place. Furthermore, water quality sensors will be installed in the inlet tank. The water and the chemicals will be moved by pumps controlled by a single-board computer (Nvidia Jetson Nano). This single board computer will work as the Data Collector and Aggregator (DCA) described in NAIADES architecture (NAIADES deliverable D2.9).

4.1 Data Collection and Aggregation

The data collector and aggregator (DCA) is the component of the laboratory in charge of collecting all the required data, process it, adapt it to NAIADES specifications and upload it to the platform. This DCA will also manage the WTP functioning, thus being possible to change treatments parameters such as coagulant and chorine dosages by manipulating dosage pumps and the filtration time by manipulating the filtration input pump.

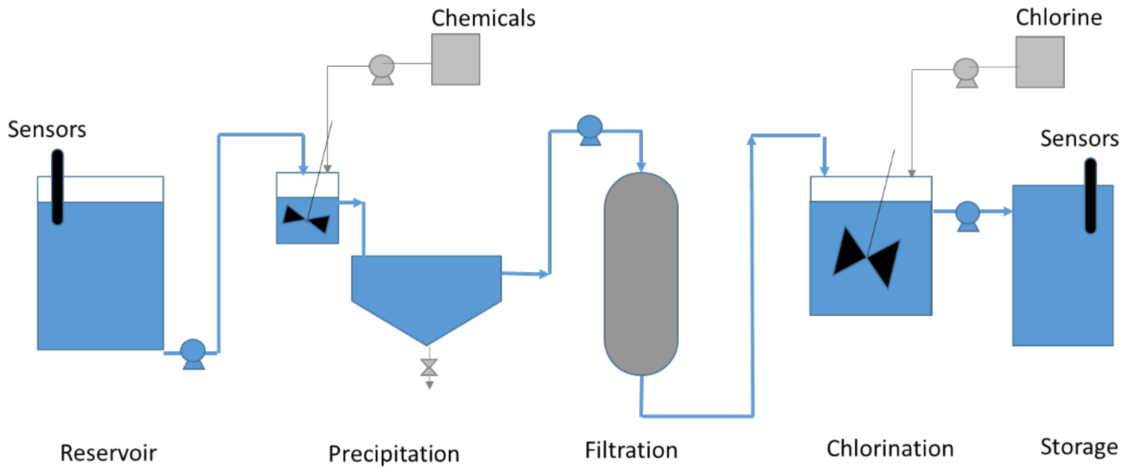

Figure 1. Laboratory drinking Water Treatment Plant. This diagram represents all the modules and the location of the sensors for the lab WTP.

4.1.1 Sensors

A total of five sensors are going to be installed in the laboratory. Turbidity (NTU), conductivity (mS/cm) and pH sensors will be installed at the inlet tank and free chlorine (mg/L) and turbidity (NTU) at the outlet tank. The outlet sensors measurements together with some laboratory measurements (nitrates and organic matter measurements) will only be used for simulation and data acquisition for simulation and validation.

4.1.2 Aggregator

The DCA oversees the aggregation of all the parameters from the laboratory that are required for the real time treatments suggestion solution.

- **Water quality measurements.** The sensors are connected to a Liquiline CM448; a digital multiparameter transmitter in charge of aggregating all the sensors measurements. This device is connected via ethernet to the DCA.
- **Treatments' parameters**. Three parameters are monitored for the Braila WTP based solution: coagulant dosages, filtration time and chorine dosage. All of them are controlled through the pumps, which are directly controlled from the DCA through USB serial commands.

- **Weather information.** The weather information is obtained from the Spanish Weather Agency AEMET. AEMET has an open API to access the last twenty-four hours hourly weather data from local stations. The weather data is not directly linked to the treatments process; therefore, this data will not be used on the modelling process since the treatments are not directly affected by the weather conditions. The weather data is indirectly related through the inlet water quality; if the quality of the inlet is known beforehand, it is possible to provide suggestions about future events, so the WTP managers can efficiently plan the necessary resources for the following days. The weather data is not necessary for the dynamical treatments suggestions tool, but it is for the Water Quality Forecast module to forecast future quality parameters at the inlet of the WTP.

Once the data is at the DCA and before being sent to NAIADES platform, it must be processed so it fulfils NAIADES requirements. Next section explains all these processes required so the laboratory data can be integrated in NAIADES platform.

4.2 NAIADES integration

The followings are common processes for NAIADES components that share data through NAIADES platform to ensure security and interoperability.

4.2.1 Common Data Models

NAIDES data models define the structure of all data shared within the platform. Any application that requires sharing data with NAIADES must follow those data structures for reading and writing data. The data models are represented in JSON format so the tools must be able to read and understand the data models; and also, to be able to create a correct formatted data model in order to submit new data to the platform.

The ls-WTP will only push data to the platform, it does not require, and will not be allowed to do any reading. The suggestions provided by the tool will be delivered to the staff of the WTP via NAIADES HMI. The laboratory is only sharing the water quality parameters measured at the inlet of the WTP and the weather data, from the closest station, on an hourly basis, thus, the required NAIADES data models are the WeatherObserved and the WaterQualityObserved. Both data models are defined using FIWARE data models with some additions of attributes.

DISY, a NAIADES partner is working on a Common Data Models Tool to help any new user to easily convert their data to NAIADES data models' format. Currently, the tool is almost ready to transform weather data from the local stations into the common NAIADES WeatherObserved data model. The laboratory will make use of this tool to transform the JSON data provided from the AEMET API to the JSON format required in NAIADES data model. This tool is created in JAVA. It will be integrated in the DCA, feeding the JSON received from the local weather station. This tool has three working programs, one per each main pilot. The laboratory will use the same program as Alicante pilot since both receive the data from the AEMET API. On the other hand, the WaterQualityObserved data model JSON format will be directly created by the DCA.

4.2.2 Digital Signature

Data signature is required to guarantee the data has not been manipulated and to allow analysis of all the data movements inside the platform.

Data signature is a service provided by one of the NAIADES modules (GT's Digital Signature Module). Each service sharing data with the platform must install the Data Signature module. Once the module is installed, the formatted water quality observations will be signed, thus, the measured water quality parameters at the inlet of the lab WTP will be a data model formatted and signed water quality observed and weather observed, respectively.

4.2.3 Tokenization

The communication with NAIADES Data Manager (DM) goes through an Access Control module that includes FIWARE Wilma and Keyrock blocks. The laboratory WTP's DCA requires to be identified by these blocks in order to be able to send and receive information. The DM will share data only to NAIADES components that have permission to access it. This identification is guaranteed through tokens. Firstly, a user and password are assigned to the DCA/pilot by the NAIADES platform managers. These credentials are used during the first communication between the DCA and the DM to obtain a token that will serve as identifier. This token is meant to be used in every communication (PUT).

4.2.4 Cloud Platform communication

The data model structure can be used to contain data from different sources. Each of these sources will create one JSON file with a unique identifier. The combination of JSON and unique id will represent an entity. The lab WTP's DCA will manage two entities:

- labWTP WaterQualityObserved, for the quality data at the inlet of the WTP
- labWTP WeatherObserved, for the weather observations

The NAIADES services that need to use these entities, should understand the nomenclature of the entities' identity and their data model structure.

The communication with the data manager will be done using NGSI-v2. It is a RESTful API via HTTP. The lab WTP uses this API through Python requests following the queries specified in the WIKI provided by the DM partner (https://gitlab.distantaccess.com/naiades/naiades-platform-poc). The entities for this use case will be created by the DM partner (UDGA) but they will be updated by the DCA.

5 Dynamical Water Treatment

This section describes the Dynamical Water Treatments Suggestions (DWTS) Tool. This tool provides a service for NAIADES platform that can be used for any end user pretending to improve the management of treatments at their drinking WTP. The version that will be integrated into NAIADES platform will be designed to optimize treatments on a drinking WTP which characteristics are similar to Braila's drinking WTP, this is, those whose treatments are coagulation, filtration and chlorination. As we mention before, this configuration is common in WTP.

The DWTS Tool aims at modelling the treatments, so the dosages and times are automatically predicted according to any water quality parameter at the inlet of the WTP. Its main purpose will be focused on water quality changes due to extreme events, such as heavy rains and droughts. Once the tool is able to accurately suggest treatments for the different events, the tool functionality will be extended to also optimize the treatments focused on normal operation.

AI techniques are going to be used for modelling and they require big data to be able to create a reliable model. Data generation is covered in the previous section (Section 4), where the laboratory WTP will simulate treatments for very different water quality parameters (correspondent to normal conditions and extreme events) that ensure the water quality parameters on the outlet are optimal (turbidity < 1 NTU and 0.5 mg/l < free chlorine < 1 mg/l according to the OMS). The DWTS Tool will use all the generated data for AI modelling.

The description of the operation of this tool can be divided into two main processes: the main and a future approach. The main process is in charge of managing real-time data and providing real-time suggestions, whereas the future approach describes the management of water quality forecasts of the inlet water to provide suggestions about future treatments.

5.1 AI modelling

The core of DWTS Tool is a machine learning algorithm that will relate a set of measurements with others. The machine learning algorithms usually requires three stages to be successfully deployed: training, validation and testing. The next subsections will describe which data is going to be used, the training process, and the validation and testing stages.

Two AI modelling processes will take place related to the tool operation processes: one for real-time and the other for future predictions.

5.1.1 Data acquisition

The required data for AI modelling is to be provided by the lab WTP, the historical data. A set containing thousands of water quality values from the inlet associated to the treatments values that guarantee the optimal outlet water quality is expected (see Table 1).

Type	Measurements	Source
Inlet Water Quality Parameters	pH	Sensor
	Turbidity	Sensor
	Conductivity	Sensor
Treatments	Coagulant dosage	Pumps' control
	Filtration time	Pumps' control
	Chlorine dosage	Pumps' control

Table 1. Data acquired for AI modelling of water treatments.

Weather data are also going to be collected as a proof of concept. For NAIADES Water Quality Prediction Services being able to provide forecasts of the inlet water quality, it is necessary to link/synchronise the input water quality parameters with the respective weather conditions. Actually, this data is not going to be correlated since the laboratory has being created inside a building and the water quality at the inlet will be mainly made at the laboratory; but collecting the weather information, all the necessary sources of information to interconnect this advanced solution (treatments suggestions forecast) are settled.

5.1.2 Training

The training phase of the modelling process requires the largest amount of data. It will take at least eighty percent of all the historical data as input of the machine learning algorithms.

Two different algorithms will be used for modelling: Support Vector Regression[1] (SVR) and a Deep Neural Network[2] (DNN). Both are supervised learning algorithms that requires a set of data containing the input and expected output data for training purposes.

SVR focuses on finding the hyperplane able to separate the input values into the specified outputs. DNN is a wide group of small functions interconnected which are adjusted (through weights) to provide outputs for any incoming data. The performance difference between both methods lies in the amount of data they require for training, their accuracy and the complexity and computation time required for training. DNN usually requires thousands of samples in order to provide reasonably good accuracies whereas the SVM can reach them with less samples. The DNN high dependency on data amount usually results in overfitting issues, this is, the trained model ends up being dependent on the incoming data; thus, the DNN data selection and preprocessing is more complex. The DNN computation time is expensive, those methods usually require very high computational power (GPU) to be trained in reasonable times, whereas SVR can be trained using CPU. Nevertheless, provided the right amount of good quality data and using the last generation GPUs, DNN can reach much better accuracies than SVR algorithms.

The SWTS tool will use these algorithms to create a model able to associate a specific combination of inlet water quality parameters to the treatments dosages and times that guarantee optimal water quality on the outlet. The hyperparameters and the Neural Network structure will be selected during this training phase; several combinations will be used for training, and the best performance one will be chosen as the final model, for SVR and DNN. Both algorithms will be tested once the historical dataset contains hundreds to thousands of values. The input parameters for both will be the same (see Table 2). Once, the training is done (and the best combination of parameters selected for each algorithm), their performance will be compared during the testing phase. It is expected to obtain better accuracy with the deep learning approach; but the SVR algorithm could be preferred if the accuracy difference is not significant since it is less computational demanding.

---

[1] Zaghloul, M. S., Hamza, R. A., Iorhemen, O. T., & Tay, J. H. (2020). Comparison of adaptive neuro-fuzzy inference systems (ANFIS) and support vector regression (SVR) for data-driven modelling of aerobic granular sludge reactors. *Journal of Environmental Chemical Engineering*, *8*(3), 103742.

[2] Al Aani, S., Bonny, T., Hasan, S. W., & Hilal, N. (2019). Can machine language and artificial intelligence revolutionize process automation for water treatment and desalination?. *Desalination*, *458*, 84-96.

Measured Variables			**Measured Variables to be predicted**		
pH	Turbidiy (Tr)	Conductivity (C)	Coagulant dosage	Filtration Time	Chlorine dosage
pH-th -5	Tr-th-20	C-th-17	+c	+t	+4cl
pH-th -5	Tr-th +5	C-th -15	+c	+10t	+1cl
pH-th -2	Tr-th +2	C-th +1	+2c	+t	+2cl
…	…	…	…	…	…

Table 2. Example of training values.

The water quality parameters at the input will have different values above and below the optimal quality (th). The treatments are the ones that will guarantee that the water quality at the outlet of the WTP will be optimal. All these values are to be generated through simulations in the laboratory WTP (through the simulator and the physical WTP laboratory – D4.3).

5.1.3 Validation and testing

The training stage returns a model that is fitted to show the best performance for the training data. The validation and testing stages' function is to improve the performance with new data and to probe the model works properly with any incoming data.

The validation stage actuates during the training stage. Twenty percent of the historical data is usually used for validation purposes. This data is used to evaluate the training model (during the training process) changing the parameters of the model, so it is not so biased by the training data.

Finally, the test stage is done once a final version of the model is created. The test stage uses completely new data to evaluate the model. The accuracy obtained with the test data will be the most representative of the model performance. The final model will be the one that probes the greater performance with the test data and it will be stored in a file to be used by the inference tool.

5.2 Model Inference

Machine learning inference refers to the use of the final model with real data. The next subsections describe how the data must be acquired and how it must be used with the model.

5.2.1 Data acquisition

The data required for inference will be taken from NAIADES platform, through the context manager. The DWTS Tool will use two different sources of data for inference: one for real time predictions and the other for future predictions. The tool will subscribe the context manager for 9 entities: 1 for current values and 8 for future predictions that are associated with the current value. These entities will contain instant values of water quality parameters: turbidity, pH and conductivity.

5.2.2 Inference

The inference process will use each of previous entities' data separately, this is, each entity will go through the final model individually. Since the algorithms to be used to create the model require the same input data, the final model selected for the tool is not changing the data inputs selection. Therefore, nine model inferences will be performed in the DWTS Tool, each of them will generate the treatments suggestions (coagulant, filtration times and chlorine) for each of the entities.

- 1 suggestion for the current water quality at the inlet. The tool uses the current data uploaded to the platform by the lab to feed the AI model.

- 8 suggestions for the future forecast (four predictions per day of the following two days). The tool will use the current future predictions uploaded to the platform by the Water Quality Prediction Tool to feed 8 instances of the AI model.

5.3 NAIADES platform integration

The integration of the DWTS Tool into NAIADES platform requires to fulfill all the NAIADES integration requirements: NAIADES data structure and signature, tokenisation, containerisation of the tool for installation and APIs and communications protocols amongst components. The last subsection is dedicated to draft a plan for a quick integration.

5.3.1 Common Data Models

The DWTS Tool will make use of the WaterQualityObserved and WaterQualityForecast data models. The modifications included in the original FIWARE WaterQualityObserved data model include the addition of parameters related to water treatments.

5.3.2 Digital Signature

The digital signature process will follow the same procedures as described in 4.2.2.

5.3.3 Tokenisation

The use of tokens for identification will follow the same procedures as described in 4.2.3. The DWTS Tool will have its own user and password associated that will be used during the first communication with the Data Manager to obtain the token. The generated token will be then used for all communications with the DM (subscriptions, put).

5.3.4 Containerisation

The Common Data Models usage, the digital signature and the tokenization are three mandatory requirements, so that any tool is able to interact within the platform. The containerisation section is about the optimal packaging of the tool, so it is easy to integrate in the Platform. Containerisation allows platform managers to quickly install any tool without worrying about compatibilities and missing libraries.

The NAIADES consortium agreed to use dockers as standard for packaging all tools. DWTS Tool is being developed using Python and it will be dockerised in three steps:

1. Environment definition. To ensure all the required Python libraries are installed.
2. Write the docker file, so it creates the environment.
3. Include the python script and the models' files.

5.3.5 Data uploading-downloading/Context Manager communications

The DWTS Tool will read the nine entities: one for the current water quality data at the inlet of the laboratory WTP and eight for the Water Quality predictions.

- The laboratory WTP inlet WaterQualityObserved. This entity is uploaded to the platform by the laboratory DCA.

- The laboratory WTP inlet WaterQualityForecast. These 8 entities are uploaded to the platform by the Water Quality Prediction Tool, a total of eight entities corresponding to the Morning, Noon, Afternoon and Night of the following two days.

The tool will return other nine entities, corresponding to the suggestions for each of the previous ones.

- The lab WTP treatments WaterQualityForecast.
- The lab WTP treatments WaterQualityForecast- Tomorrow Morning.
- The lab WTP treatments WaterQualityForecast- Tomorrow Noon.
- The lab WTP treatments WaterQualityForecast- Tomorrow Afternoon.
- The lab WTP treatments WaterQualityForecast- Tomorrow Night.
- And another four for the day after tomorrow.

The other NAIADES services that need to use these entities should understand the nomenclature of the entity's identity and the data model structure.

The communication with the data manager will be done using NGSI-v2. It is a RESTful API via HTTP. The Water Quality Prediction Tool uses this API through Python requests following the queries specified in the WIKI provided by the DM partner (https://gitlab.distantaccess.com/naiades/naiades-platform-poc). The entities for this use case will be created by the DM partner (UDGA) but they will be updated by the data providers.

- The lab WTP inlet WaterQualityObserved – by AIMEN, from laboratory WTP's DCA.
- The eight lab WTP inlet WaterQualityForecast – by AIMEN, from Water Quality Prediction service.
- All the lab WTP treatments WaterQualityForecast entities – by AIMEN, from this service.

Data reading will be done through subscriptions to the real time DM module. This means, the DM will send data to the tool each time a new value of each entity is uploaded to the platform.

5.3.6 Integration

This section shortly describes the integration planning for the DWTS Tool. The specifications of the infrastructure required to be able to run the DWTS Tool will depend on the final machine learning algorithm chosen. The simplest algorithm (SVR) can be run with CPU computational power and at least 8GB of RAM whereas the DNN solution will require GPU (similar to NVidia GeForce GTX 1080 or superior) to perform smoothly. The integration is expected to take place in two phases:

1. First model deployment. Once the first model has being created three steps will be taken:
    a. The tool will be validated locally reading data from NAIADES development platform. This phase should serve to validate the correct use of the entities, the data signature and the interaction with the DM using ngsi-v2 API.
    b. The tool will be dockerised and sent to the NAIADES production platform (charged by SIMAVI) for installation. During this first interaction with the production platform, all the problems regarding the tool integration in the production platform will be attended.
    c. Tool performance evaluation in the production platform.
2. Model improvement.
    a. The model will be constantly trained with new data. Each time when a better performance model is trained, it will be shared with SIMAVI for integration.
    b. Each new model will be also evaluated in the production platform.

This plan relies on the availability of data. The models could be improved only if the amount of data increases.

6 Next Steps

This section describes the current and future objectives. Currently, there is no data available to create any model for treatments suggestions. The data required for modelling will be provided by NAIADES lab WTP pilot and since the lab is a countermeasure plan since the real WTP cannot provide treatments data, the obtainment of data for this tool is being delayed. Data from the simulator is expected soon, but data generated by the physical lab for model optimization and validation is expected near the end of this task period. Once there is enough data available for modelling the next steps will be followed.

- **Algorithm selection**. Both algorithms (SVR and DNN) will be trained, and their performance evaluated. The algorithm with higher accuracy model will be chosen. If the accuracies difference is not significant, the less computational demanding will be chosen.
- **Periodical model upgrade.** The model will be retrained when new data is available (in the order of hundred new samples). It is expected to have periodical model upgrades with new real data until October 2021.
- Once the model performance is optimal, if there is enough time, the advanced solution for **future treatments suggestions** will be demonstrated.

7 Correlation between weather nowcasting/forecasting and wastewater treatment process control

Water Treatment processes' control is not directly linked to weather conditions. The processes will differ depending on the water quality measured at the inlet of the dWTP; and it is the quality of the inlet water the one directly linked to the weather conditions, since rain or drought events will produce changes in water quality parameters such as Turbidity increments. Therefore, what it is mentioned in this deliverable is the possibility to combine the Dynamical Treatments Suggestion service with NAIADES water quality forecast service (D5.9). NAIADES water quality forecast service can provide future values of the water inlet quality; and the Dynamical Treatments Suggestions service can suggest treatments for those future values. Therefore, the weather information is required by the Water Quality Forecast service.

8 Conclusions

This report includes a description of the elements required to develop, implement and validate the Dynamical Water Treatments Suggestion Tool. A tool intended for automatically predicting the treatments dosages and functioning times of a WTP when the water quality at its inlet changes drastically due to extreme events such as heavy raining.

The first approach to develop this tool required historical data (water quality parameters at the inlet and water treatments used) from a real WTP. However, this first approach was changed due to the impossibility to get treatments information from WTP for confidentiality issues. Therefore, a counter measurement plan was activated. This plan required anticipating the creation of a scaled version of a WTP in a laboratory where all the extreme events could be simulated (through SW simulators and on the physical WTP lab – see D4.3). The application of this plan has delayed the generation of data, which affected the development of the DWTS Tool. Currently, the laboratory is being constructed, and the first data generated is expected in one month. Furthermore, data generation will take time since the operation of the laboratory WTP will take a couple of days. As a result, it is not expected to have enough data for data modelling until mid-2021.

In the Section 4 of this document describes the technical part of the laboratory that is required for data collection and communication with NAIADES platform. It explains all data involved, the tools and processes required to adapt to NAIADES requirements, such as data modelling, digital signature, tokenization and communications with the cloud platform. Section 5 focuses on the DWTS Tool itself, explaining its internal functioning (machine learning algorithms based), all the required data from NAIADES, and the processes to adapt to the tool for NAIADES requirements, which includes, data modelling, digital signature, tokenization, dockerisation and communications with the data manager of the platform.

Finally, the integration plan of the DWTS Tool and the next steps are explained. Both of them are completely dependent on data availability, which is the highest risk to overcome to succeed in the implementation of this tool.